DOE DATA DAYS 2022

APPFL

Argonne NATIONAL LABORATORY

APPFL: OPEN-SOURCE FRAMEWORK FOR PRIVACY-PRESERVING FEDERATED LEARNING

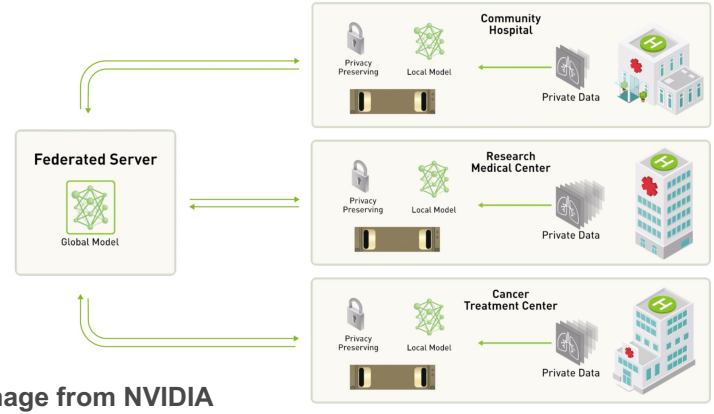


Image from NVIDIA

KIBAEK KIM
Computational Mathematician
Mathematics and Computer Science
Argonne National Laboratory

MINSEOK RYU
YOUNGDAE KIM
Postdoctoral Appointees
Mathematics and Computer Science
Argonne National Laboratory

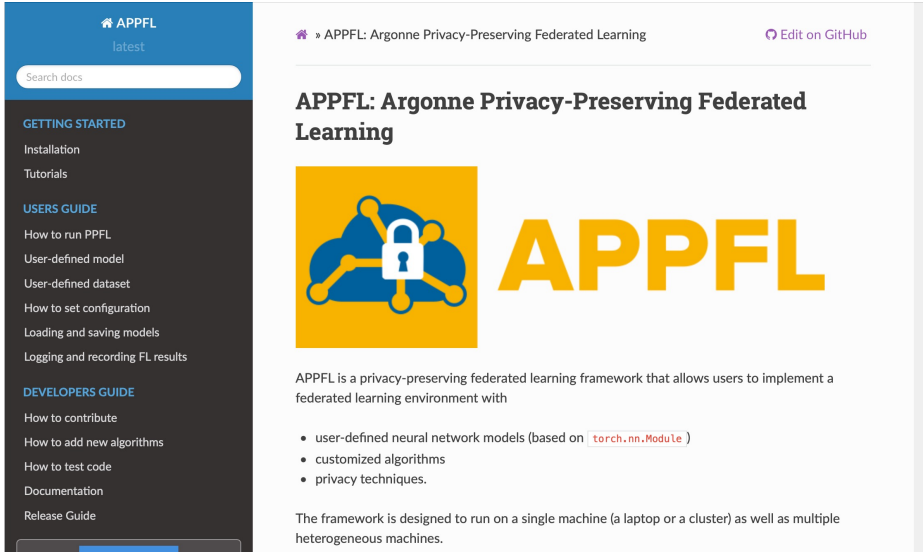
RAVI MADDURI
Computer Scientist
Data Science Learning
Argonne National Laboratory

June 2, 2022

APPFL: ARGONNE PRIVACY-PRESERVING FEDERATED LEARNING

Open-source PPFL framework

- Motivations
  - Distributed data from multiple institutions (e.g., Argonne, UChicago, etc.)
  - Avoid data transfer to a server
  - Potentially sensitive/private data
- Science applications
  - Biomedical data
  - Smart meters deployed in grid
  - Experimental facilities (e.g., APS X-ray beamlines)
  - National security (e.g., critical infrastructure)



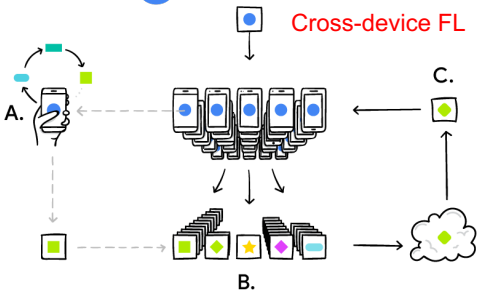
Argonne NATIONAL LABORATORY

FEDERATED LEARNING

- Machine learning without centralizing training data
  - No direct data sharing or storing
- More benefits
  - Learning a global/shared model
  - Utilizing a localized model at each client side
  - Personalization
- Two settings:
  - Cross-device FL (1000s and 1Ms of small devices)
  - Cross-silo FL (a few large data repositories)

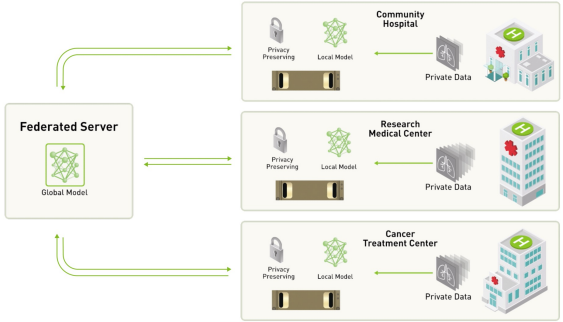
Google

Cross-device FL



Your phone personalizes the model locally, based on your usage (A). Many users' updates are aggregated (B) to form a consensus change (C) to the shared model, after which the procedure is repeated. (image from Google)

Cross-silo FL

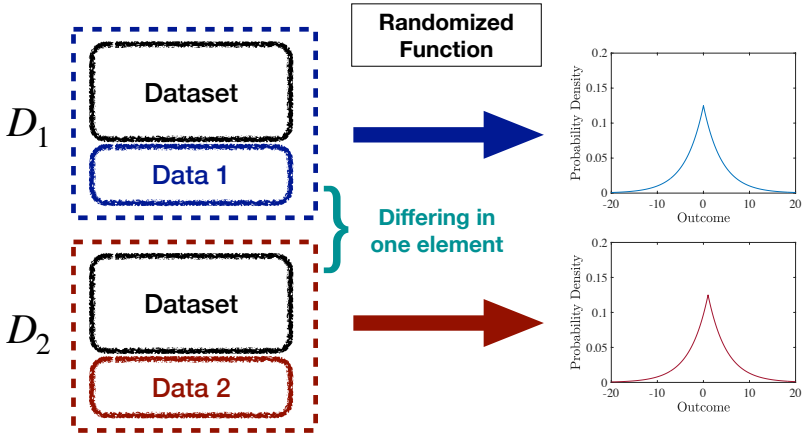


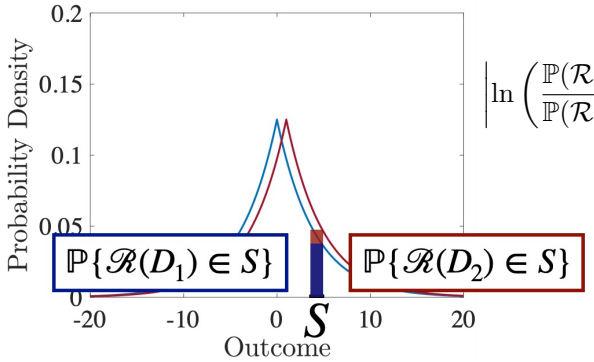
Federated learning on decentralized medical datasets (image from NVIDIA)

Argonne NATIONAL LABORATORY

PRIVACY-PRESERVING TECHNIQUES

- **Some techniques in FL**
  - Homomorphic encryption: limited to certain operations
  - Secure multi-party computation: computationally expensive
  - Differential privacy: potential accuracy loss
- **Differential Privacy**
  - The two outcomes are indistinguishable for all D1 and D2 which differ in one individual's data.

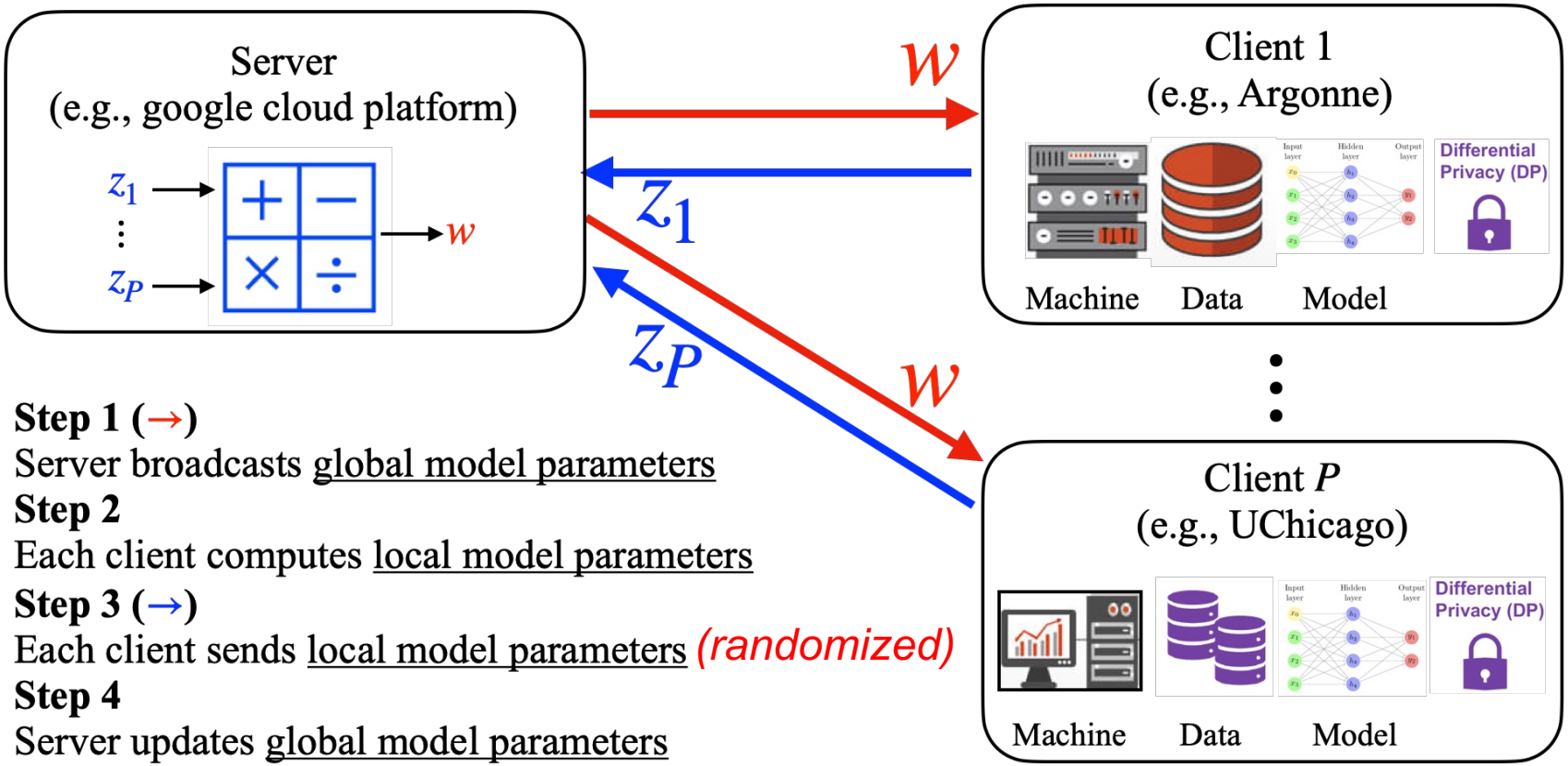




$$\left|\ln\left(\frac{\mathbb{P}(\mathcal{R}(D_1) \in S)}{\mathbb{P}(\mathcal{R}(D_2) \in S)}\right)\right| \leq \epsilon$$

Argonne NATIONAL LABORATORY

OVERVIEW: APPFL FRAMEWORK



Step 1 (→)
Server broadcasts global model parameters
Step 2
Each client computes local model parameters
Step 3 (→)
Each client sends local model parameters *(randomized)*
Step 4
Server updates global model parameters

MAJOR COMPONENTS OF APPFL

- **Training algorithms:**
  - IIADMM, FedAvg [McMahan et al., 2017], ICEADMM [Zhou and Li, 2021]
  - Any user-defined FL algorithms can be added.
- **Differential privacy**:
  - Random perturbation with Laplacian noises [Dwork et al., 2006]
  - More advanced schemes can be added.
- **Communication protocols**:
  - gRPC: communication between multiple platforms and languages
  - MPI: efficient communication in a cluster environment
- **User-defined model and data:**
  - Inherits PyTorch's neural network module, torch.nn.Module
  - Dataset class that inherits the PyTorch's Dataset

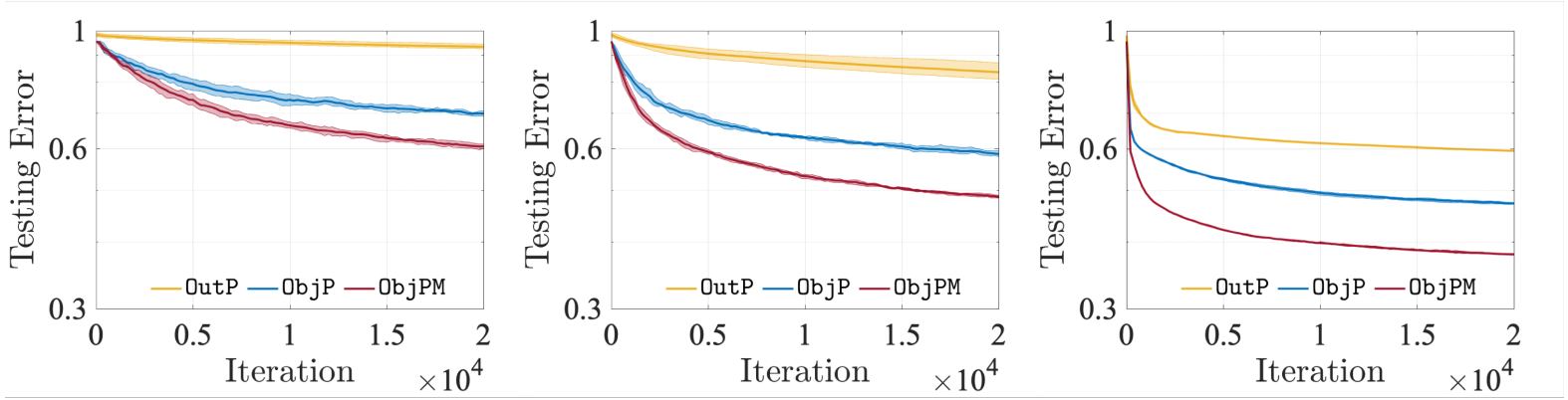
ADVANCED PPFL ALGORITHMS

Implementation of novel training algorithms

- (state-of-the-art) OutP: Inexact ADMM (IADMM) + output perturbation
- (APPFL) ObjP: IADMM + objective perturbation
- **(APPFL) ObjPM:** IADMM + objective perturbation + multiple local updates

Stronger privacy
Weaker learning

Weaker privacy
Stronger learning

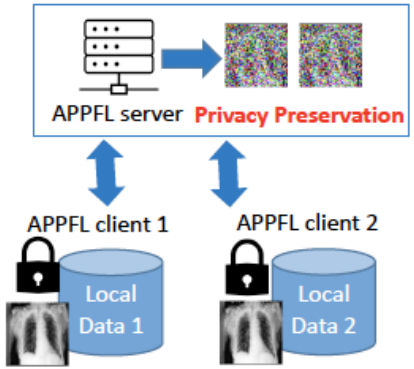


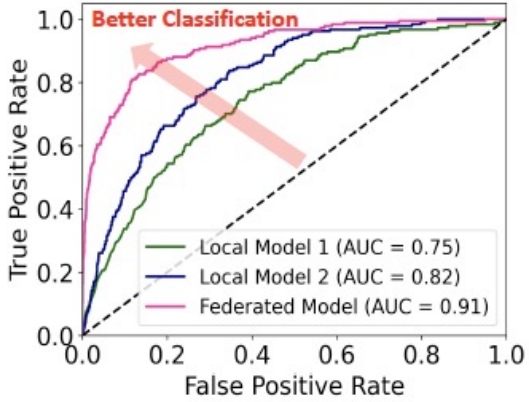
Argonne NATIONAL LABORATORY

USE CASE

Chest X-ray classification for COVID-19 cases

- FL can produce more accurate model, compared to the models trained on local datasets.
- DP is applied to protect chest X-ray data from reverse-engineering the model gradients communicated during training.
- Collaborations with UChicago Medical School and Broad Institute

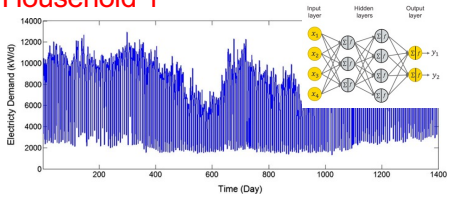




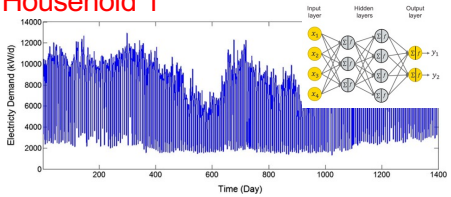
USE CASE

Federated load forecasting for electric distribution system

Household 1

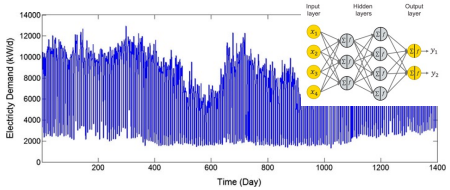


Household 1



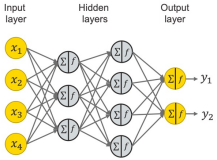
⋮

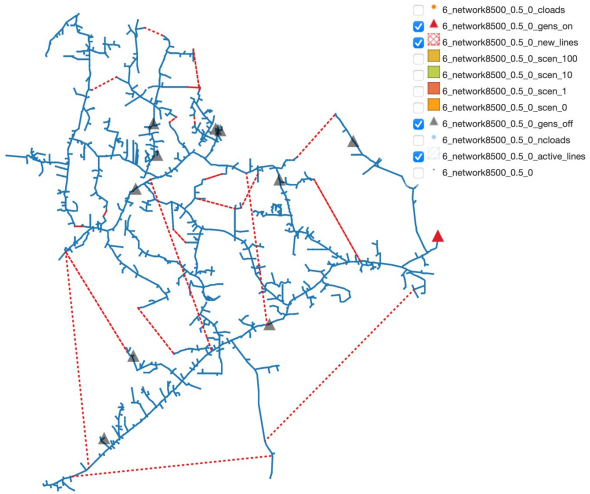
Household *N*



Parameters with differential privacy

APPFL





- Electricity consumption at household level
- Distributed data at a large number of devices
- Privacy concern

Argonne NATIONAL LABORATORY

CONCLUDING REMARKS

- APPFL: open-source Python package with support of any Pytorch models
  - The package has been released (v0.2.0).
- Any user-defined ML model can be trained on decentralized data while ensuring data privacy.
- Customized PPFL algorithms can be easily implemented.
- Applications: national security, smart grid (Fig. 1), biomedicine, experiments (Fig. 2), etc.
- Collaborations and contributions are welcome!

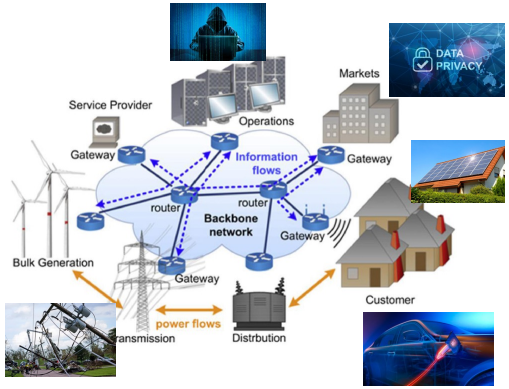


Fig 1. Network architecture in smart grid (modified from https://doi.org/10.1016/j.comnet.2012.12.017)

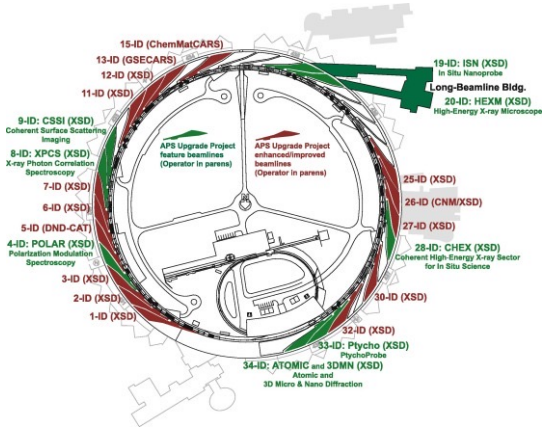


Fig 2. Multiple experimental devices at APS

Argonne NATIONAL LABORATORY

REFERENCES

- Minseok Ryu and Kibaek Kim. “Differentially Private Federated Learning via Inexact ADMM with Multiple Local Updates” *arXiv preprint arXiv:2202.09409*, 2022.
- Minseok Ryu, Youngdae Kim, Kibaek Kim, and Ravi Madduri. “APPFL: Open-Source Software Framework for Privacy-Preserving Federated Learning” *arXiv preprint arXiv:2202.03672. (accepted to IPDPS 2022)*, 2022
- Minseok Ryu and Kibaek Kim. “A Privacy-Preserving Distributed Control of Optimal Power Flow” *IEEE Transactions on Power Systems 37(3)*, 2022
- https://github.com/APPFL

ACKNOWLEDGEMENT

Funding supports and Contributions

- DOE-ASCR PALISADE-X Project
- DOE Early Career Research Program (on data-driven optimization)

- Minseok Ryu (ANL)
- Youngdae Kim (former postdoc at ANL)
- Ravi Madduri (ANL)
- Jordan Fuhrman (UChicago)
- Nick Dodd (ASU)

THANK YOU